

基于内容标注的三角引用动机研究方法探析^{*}

■ 刘运梅^{1,2} 张帅^{1,2} 司湘云^{1,2} 马费成^{1,2}

¹ 武汉大学信息管理学院 武汉 430072 ² 武汉大学信息资源研究中心 武汉 430072

摘 要: [目的/意义] 文献三角引用是一种融合直接引用、共被引与耦合关系为一体的引用结构,具有较高的研究价值与应用拓展意义。为从深层次认识和挖掘三角引用现象,尝试通过内容标注的方式深入挖掘其中隐藏的引用动机。[方法/过程] 首先通过整理已有文献,构建一个包含功能性引用动机和情感性引用动机的动机标注框架;其次,构建三角引用数据的获取步骤,并在 CNKI 数据库中获取 9 442 条三角引用关系数据;最后,专业人员根据预设的动机框架,为每条三角引用数据中的三种引用关系标注相应的引用动机。[结果/结论] 实验结果发现:三角引用结构中的施引行为倾向于包含多个功能性引用动机,而情感性引用动机则多具有唯一性;B→A 与 C→A 两种引用关系的引用动机分布基本一致,但与 C→B 存在差异;原始文献 A 一般是相关研究主题、领域或学科比较重要的、高影响力文献,倾向于提供一些新颖的、开创性的或基础的思想、概念、观点或方法。

关键词: 三角引用现象 引用动机 引用功能 引用情感 内容标注

分类号: G250

DOI: 10.13266/j.issn.0252-3116.2021.10.006

1 引言

文献直接引用、文献耦合^[1]与文献共被引^[2]是目前最常见的 3 种文献引用关系。其中,直接引用是文献间最基础、简单的引用关系;文献耦合与共被引分别在 1963 年^[1]和 1973 年^[2]被提出,广泛应用在研究前沿探测、信息检索、知识结构分析等领域。其中,共被引分析具有时滞性,适用于分析出版时间较久的文献集;此外,文献间的共被引关系及共被引强度随着时间推进具有变化性,有助于学科发展态势和主题演变分析,但无法自动对筛选出的文献进行主题描述,需要人工辅助解读。相比之下,在精确度和识别结果数量上,耦合分析优于共被引分析;在时间维度上,耦合分析探测速度也优于共被引分析;但由于参考文献的固定性,文献间的耦合关系形成后不会随时间变化,不适用于主题演变分析^[3]。因此,共被引分析与耦合分析都能将无外部关联的文献客观地联系起来,揭示一组文献间错综复杂的交叉关系和联系程度,但这两种对偶关系所体现的引证规律和结构关系的角度各不相同、各

有优势^[4-5]。

文献三角引用则将文献直接引用、共被引与耦合关系融合到一起,形成了另外一种特殊的引用关系,其定义如下:若文献 A 与文献 B 被文献 C 引用,同时文献 A 与 B 之间也存在引用关系,那么 A、B、C 三者之间就建立了三角引用关系。文献三角引用是一种多元的引用现象,其中既有直接引用关系,又同时具有共被引关系(A 与 B 被 C 引用)和耦合关系(B 与 C 都引用了 A)。

三角引用结构中包含了 3 种文献:A、B、C,和 3 种引用-被引关系:文献 B 引用 A(B→A)、文献 C 引用 A(C→A)、文献 C 引用 B(C→B)。从引用时间看,文献 B 引用 A 是最早发生的引用关系,其后,文献 C 是在文献 A 与 B 正式发表后才对 A、B 进行引用。因此,3 种文献的发表时间先后顺序是:A、B、C;3 种引用关系发生的时间顺序是:B→A 先发生、接着 C→B 与 C→A 同时发生。根据 3 种文献在施引行为上的主体方和客体方、在时间轴的分布位置,将文献 A 取名为“原始文献”,文献 B 称为“中介文献”,文献 C 称为“追随文

^{*} 本文系国家自然科学基金重点国际(地区)合作研究项目“大数据环境下的知识组织与服务创新研究”(项目编号:71420107026)研究成果之一。

作者简介: 刘运梅(ORCID:0000-0002-8921-0826),博士研究生;张帅(ORCID:0000-0002-5792-877X),博士研究生,通讯作者,E-mail:zs09053@163.com;司湘云(ORCID:0000-0003-1367-0532),博士研究生;马费成(ORCID:0000-0003-0187-013),教授,博士生导师。

收稿日期:2020-12-10 **修回日期:**2021-02-26 **本文起止页码:**48-55 **本文责任编辑:**易飞

献”。引用结构如图 1 所示：

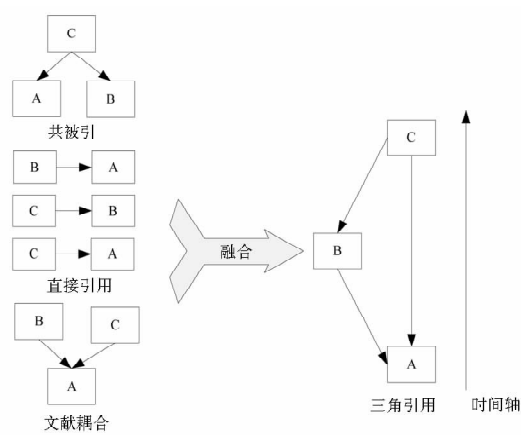


图 1 文献三角引用结构

在前期研究中,我们发现上述的三角引用结构广泛存在于实际的文献引用网络中,并且从文献特征分析、引文内容特征分析等多维角度发现,三角引用结构中的文献 A、B、C 各有不同的角色和价值,B→A、C→A、C→B 这 3 种引用关系也各有不同的引用特征。通常情况下,原始文献 A 是被引数量最多、被引用强度最大、被引用位置最靠前的文献;中介文献 B 是三角引用机制中关键的一环,起到联通作用;追随文献 C 则是三角引用结构中最活跃的施引角色,主导了三角引用关系的产生。其次,在 3 种引用关系中,C→A 的引用特征明显不同于其他两种直接引用(B→A、C→B),在文献 A 与 C 之间普遍存在一种“间接引用”机制。若追随文献 C 与原始文献 A 存在语言差异、文献类型差异或学科差异,但与中介文献 B 处于同一语言环境、同一种文献类型、同一学科或同一学者(团队)研究课题,在这种情境下,追随文献 C 为了引用文献 A,往往会出现惰性引用动机,部分转引、甚至直接全文转引 B 中引用 A 的引文内容,从而导致 C→A 的引文内容与 B→A 出现高度一致。

综上,三角引用现象是一种常见的、特殊的文献引用关系,其中蕴含着重要的、复杂的“引用-被引用”内涵和机理,其中的 3 个引用关系 B→A、C→A、C→B 并不能被简单等同于一般的直接引用关系,具有复杂的引用动机与情境。因此,有必要深入挖掘三角引用现象的引用行为与动机,一方面,以更好地认识三角引用这一新的结构与现象;另一方面,以发现国内科学文献之间的引用动机特点与规律。

引用动机是指施引作者对参考文献的引用目的,通常可被视为作者理性的心理行为,是一个主观而非客观的变量^[6-7]。在前期研究中,无论是通过文献外

部特征分析,还是引文内容特征挖掘的三角引用现象研究,都是建立在文献外部特征或文献语义特征所表征的规律之上,极有可能与研究者写作时的真实引用动机存在出入,带有一定间接性。因此,本文尝试根据现有引用动机的相关研究,构建学者引用动机的分类框架;其次,获取一定数量的三角引用关系数据,组织相关学科的科研人员根据具体的引文内容和构建的动机分类模型,对采集到的三角引用关系进行引用动机分类,探究三角引用现象中的引用动机与影响因素等,以期揭示三角引用现象隐含的驱动因素、引用规律,进一步丰富引文分析方法和科学评价理论。

2 引用动机相关研究

在中国科学研究领域,一些评价体系已经对引文数据提出了要求^[8],例如,中国国家自然科学奖评价指标明确规定:“主要学术思想和观点被认可是指他人在正式发表的论文、专著中正面引用完成人提出的学术思想、观点或方法的情况。”因此,文献的引用动机十分重要,其既能结合引文内容等特征,发现学科领域之间的引用规律,还能为基于引文的学术评价提供参考。目前,引用动机的相关研究可分为 3 类:从理论上对引用动机归类;直接通过访谈、调查问卷获取作者的真实引用动机;基于引文内容识别作者的引用动机,包括人工标注、自动标注。

2.1 引用动机分类

关于引用动机的分类目前还没有统一的标准,不同的学者持有不同的观点。例如:E. Garfield 根据被引文献在施引文献中的位置、内容、形式等,将引用动机较为系统地划分为 15 种经典动机,包括向开拓者致敬、提供背景资料、充实性的声明、对他人工作的否定等,这是最早提出深入研究引文动机的提议^[9];M. J. Moravcsik 等的研究也具有一定的代表性,他们采用二分法从 4 个维度建立了引用性质分类表——概念性或操作性引用、陈述性或敷衍性引用、扩展或继承性引用、质疑或否定性引用^[10];P. Vinkler 则将作者的多种引用动机归为两大类:专业动机和关系动机,专业动机是指由于理论或实践上的内容联系导致作者的引用行为,而关系动机是指作者为了和学术共同体建立起社会联系而进行的引用^[11];类似地,朱大明也基于引文原理和影响引用动机的各种因素,将参考文献的引用动机分为学术性动机和非学术性动机,并将学术性动机进一步划分出 10 种小类^[12]。

2.2 访谈与问卷调查法

早期受全文本研究的技术限制,学者们大多通过

调查问卷或访谈来开展引用动机研究。调查的基本方式有两种:①直接询问作者撰写论文与施引文献之间的关联方式与程度。例如:T. Rong 等询问了 99 位生物学家和心理学家的引用动机,并请每位作者对本人的每篇引文给予重要性判断,发现大部分的引用动机是提供研究背景信息^[13];史亚莉等采用扎根理论方法,调查引证过程中科研用户的数据认知行为特征,确定了引证过程中科研用户数据认知行为的范畴体系^[14];宋维翔则通过邮件访谈的方式访问了若干睡美人文献和王子文献的作者,通过他们对 3 个引用动机问题的回复,发现形成睡美人文献的原因主要有作者的谦虚性、团队的不稳定性等因素^[15]。②首先预设好一个引用动机框架,让施引作者对其论文中的引文进行动机归类。例如:T. A. Brooks 首先将引用动机分为 7 种类型,再对爱荷华大学的科研人员开展调查,让他们指出其发表论文中每一条引文的引用动机属于哪一类或几类,实验发现作者在引用一条文献时可能存在多种动机^[16];M. Singson 等构建了一份由人口统计资料和 23 项引文信任和引文来源权威声明组成的预先定义的问卷,再邀请 100 名高校教师填写,实验结果发现引用行为是复杂且多面的,有部分研究人员的引用行为会受到社会关系影响^[17];张敏等提出研究假设和模型,采用情境实验结合问卷访谈的实证方法,探究了科研人员引用行为的影响因素与内在认知过程^[18];邱均平等也提出了针对科研人员引用行为的影响因素模型,包括内在和外在引用动机两个方面,并通过问卷结果来分析各引用动机之间的相互影响关系^[19]。

2.3 基于引文内容的动机标注

除对学者自身展开引用动机调查外,其他研究还通过引文内容信息来标注引用动机。其中,人工标注一般是由专业的标注人员根据预先的动机分类标准进行标注,如:丁文姚等采用内容分析法从 9 个维度对样本论文的科学数据引用行为进行数据编码,应用统计学方法描述图情领域期刊论文科学数据引用特征,并探索不同维度特征间的关联关系^[20];S. Teufel 等将引用动机分为不足、肯定、对比和中立 4 个类别,并在此基础上继续细分为 12 个标准,另外又将这 12 个标准归到三大类别——负面、中性、正面,然后由 3 名标注人员按照上述分类标准对 26 篇文章的 548 次引用进行动机标注,发现按照这类标准的标注结果一致性较高^[21]。

此外,随着自然语言处理、机器学习技术成熟,越来越多的学者尝试利用计算机标注引用动机:S.

Teufel 等利用内容语义结构与特征词对引用情感分类,提出了一种利用动词线索词的引用内容自动分类方法^[21];刘盛博等通过自动句法标注寻找引用相关句子的主语,并根据主语和 Teufel 动词线索词的搭配关系对引用内容自动分类,发现在 BMC-Bioinformatics 期刊中,62.88% 的引用是中性的,正面引用达到了 33.59%,而负面引用只有 3.53%^[22];M. Roman 等认为利用深度神经网络算法对引文进行深入分析,可以揭示有趣的发现,他们利用引文上下文数据集对 1 000 万条引文上下文进行标注,发现 BERT 嵌入的准确率最高^[23];M. Y. Wang 等提出了一种结合引文句法特征与语境特征的机器学习框架,通过检查引文的句法和上下文信息来区分重要和不重要引文,实验结果表明该分类框架能够获得较好的分类性能^[24]。

综上所述,科学研究中的引用动机蕴含着丰富的信息,通过访谈或调查问卷方式可以直接反映作者对参考文献的引用动机,但实验数据少且操作复杂,容易忽略其他重要的引文内容信息。而通过机器标注方法的引用动机适用于识别具体数据集的引用动机,并且带有一定的间接性、随机性。因此,本文在基于现有文献研究的基础上,构建关于三角引用结构的引用动机标注框架,并邀请专业人员根据具体的引文内容信息,对采集到的真实三角引用关系数据进行动机标注,并对统计与调查结果进行分析、归纳,尝试挖掘三角引用现象的真实引用情境与动机,进一步揭开三角引用现象这一多元、复杂引用结构的面纱。

3 引用动机分类框架构建

本文对目前国内外有关引用动机的研究成果进行梳理、总结、分类,构建引用动机分类框架。其中,根据引用动机的性质,将其分为功能性引用动机和情感性引用动机两大类,并在此基础上进一步对两种类别继续细化、分类,具体分类框架与参考文献来源见表 1。

4 三角引用数据采集

在数据采集阶段,首先需要建立三角引用关系的获取步骤。本文以原始文献 A 入手,寻找中介文献 B 和追随文献 C,来确定以文献 A 为原始文献的三角引用数据,步骤如下:首先,采集引用文献 A 的所有施引文献,得到中介文献集合 $\{B_0, B_1, B_2, \dots, B_i, \dots\}$,即多个“B→A”的引用关系对。其次,分别采集中介文献集合中每个文献 B 的施引文献。最后,获取 A 的施引文献与 B_i 的施引文献中相同的文献,所得到的相同文献就

表 1 引用动机分类

性质	引用动机分类		参考来源
功能性引用动机	参考文献提供了背景信息	参考文献是本研究领域近期发表的代表性文献,引用参考文献说明现有文献未涉及该论题,阐明该论题的新颖性或前沿性	文献[12, 25–26]
		引用参考文献的相关事实或数据,说明该研究主题的重要性或必要性	文献[12, 25–26]
	采用了参考文献中提出的新概念、新观点或新方法等		文献[19, 27]
	为综述、评论以前的文献而引用		文献[27]
	验证参考文献中的数据、公式、方法、工具等		文献[27]
	引用图表等		文献[28–29]
	参考文献提供了用于比较的信息或数据	引述他人的理论、方法、结果或结论,与作者自己的文献进行对比分析	文献[26, 30]
		作者对参考文献的成果进行总结、评价、否定等	文献[26, 30]
	在结论或讨论中使用参考文献总结		文献[29, 31]
	为了自引,使研究具有延续性而引用自己的文献		文献[32–34]
情感性引用动机	举例引用,为研究成果提出实例而引用文献		文献[33–34]
	正面情感	对该领域的开拓者或创始人表示尊重	文献[25]
		对参考文献或其作者的工作表示肯定	文献[25]
	中性情感	用于文献综述	文献[11, 35]
		不带情感引述参考文献的概念、观点	文献[35]
		客观描述参考文献的数据和事件	文献[35]
	负面情感	指出参考文献中的不足之处,否定参考文献中的观点	文献[25]
		对参考文献作者的优先权提出异议	文献[25]

是追随文献集合 C_i , 即“ $C_i \rightarrow A$ ”, 同时“ $C_i \rightarrow B_i$ ”。那么, 文献 A 、 B_i 与集合 C_i 中的每个文献就组成了三角引用关系。

本文选择中国学术期刊网络出版总库 CNKI 作为数据来源。考虑到在 CNKI 数据库中期刊论文具有较规范的引文体系和记录, 选用 2015–2020 年“图书情报与数字图书馆”类目下被引频次最高的前 20 篇期刊论文作为“原始文献”。并参考以上三角引用数据的获取步骤, 以这 20 篇高被引期刊论文入手, 采集三角引用数据样本, 共发现了 10 875 条三角引用关系, 数据采集时间为 2020 年 1 月 18 日至 2 月 23 日, 数据爬取工具为 Python。表 2 列出了关于原始文献 A 的三角引用关系统计量。

以 20 篇高被引期刊论文作为原始文献入手, 本文共发现了 10 875 条三角引用关系。如表 2 所示, 在这 10 875 条三角引用关系中, 原始文献 A 共 20 篇, 中介文献 B 共 2 505 篇, 追随文献 C 共 4 533 篇。为了提取三角引用关系中的具体引用位置和引文内容数据, 利用 Python 爬虫程序获取 $B \rightarrow A$ 、 $C \rightarrow A$ 、 $C \rightarrow B$ 三种引用关系的施引文献全文信息, 即 2 505 篇文献 B 和 4 533 篇文献 C 的 XML 格式全文数据。其中, 有 648 篇文献的全文中仅有参考文献信息, 在原文未标注具体的引用位置, 因此无法获取这些文献的引文位置和引文内容数据, 本文便将这些文献所在引用关系的相关数据

表 2 关于原始文献 A 的三角引用关系统计数据

原始文献序号	原始文献被引数量/篇	中介文献数量/篇	追随文献数量/篇	三角引用个数
1	1 486	466	848	2 414
2	666	219	359	882
3	631	66	151	186
4	528	135	298	569
5	512	153	248	528
6	486	152	236	588
7	484	56	77	103
8	460	95	183	313
9	459	103	187	313
10	450	88	157	244
11	442	10	11	12
12	436	91	181	362
13	403	117	227	605
14	394	96	169	332
15	392	102	175	381
16	384	140	296	1 083
17	381	112	185	453
18	377	70	105	147
19	376	111	194	469
20	374	123	246	891
总计	10 121	2 505	4 533	10 875

填为空值, 共剩余 9 442 条可供实验分析的三角引用关系数据。

邀请两名图书情报学领域专业的科研人员参照前

ChinaXiv:202304.00641v1

文预设的引用动机分类框架和每条三角引用数据中的 $B \rightarrow A$ 、 $C \rightarrow A$ 、 $C \rightarrow B$ 三个引用文本,进行动机标注。其中,标注原则是:①每条“引用-被引”关系必须同时包含功能性引用与情感性引用中的引用动机;②:每条“引用-被引”关系可对应功能性引用或情感性引用中的一个或多个引用动机。为了保证标注实验的可靠性,随机选取 10% (944 条) 的三角引用关系数据作为对两名标注人员的测试,两位标注人员标注结果的 Cohen kappa 系数为 0.75,表明他们标注结果之间的一致

性较高^[36],其他标注分歧通过讨论得到了解决。然后,剩下的 90% 的三角引用数据由两位编码人员平均完成。

5 三角引用动机的调查结果分析

根据标注结果,分别统计三角引用结构中三种引用关系的功能性引用动机与情感性引用动机分布情况,并计算 18 种引用动机在 9 442 条三角引用关系中对应的覆盖程度,统计结果见表 3 与表 4。

表 3 三角引用关系的功能性引用动机标注结果

功能性引用动机分类	数量 ($B \rightarrow A$)	覆盖率/% ($B \rightarrow A$)	数量 ($C \rightarrow A$)	覆盖率/% ($C \rightarrow A$)	数量 ($C \rightarrow B$)	覆盖率/% ($C \rightarrow B$)
参考文献是本研究领域近期发表的代表性文献,引用参考文献说明现有文献未涉及该论题,阐明该论题的新颖性或前沿性	729	7.72	822	8.71	982	10.40
引用参考文献的相关事实或数据,说明该研究主题的重要性或必要性	1 155	12.23	1 213	12.85	786	8.32
采用了参考文献中提出的新概念、新观点或新方法等	1 481	15.69	1 292	13.68	610	6.46
为综述、评论以前的文献而引用	5 954	63.06	6 146	65.09	7 259	76.88
验证参考文献中的数据、公式、方法、工具等	269	2.85	175	1.85	520	5.51
引用图表等	388	4.11	545	5.77	452	4.79
引述他人的理论、方法、结果或结论,与作者自身进行对比分析	320	3.39	268	2.84	426	4.51
作者对参考文献的成果进行总结、评价、否定等	951	10.07	1 039	11.00	1 129	11.96
在结论或讨论中使用参考文献总结	812	8.60	783	8.29	906	9.60
为了自引,使研究具有延续性而引用自己的文献	956	10.12	71	0.75	233	2.47
举例引用,为研究成果提出实例而引用文献	556	5.89	421	4.46	350	3.71

表 4 三角引用关系的情感性引用动机标注结果

情感性引用动机分类	数量 ($B \rightarrow A$)	覆盖率% ($B \rightarrow A$)	数量 ($C \rightarrow A$)	覆盖率% ($C \rightarrow A$)	数量 ($C \rightarrow B$)	覆盖率% ($C \rightarrow B$)
对该领域的开拓者或创始人表示尊重	872	9.24	919	9.73	252	2.67
对参考文献或其作者的工作表示肯定	1 940	20.55	1 707	18.08	986	10.44
用于文献综述	3 945	41.78	3 232	34.23	4 950	52.43
不带情感引述参考文献的概念、观点	2 428	25.71	2 791	29.56	2 692	28.51
客观描述参考文献的数据和事件	1 214	12.86	1 322	14.00	1 042	11.04
指出参考文献中的不足之处,否定参考文献中的观点	302	3.20	257	2.72	354	3.75
对参考文献作者的优先权提出异议	68	0.72	83	0.88	74	0.78

通过对上文的标注与统计结果,可以发现:
(1)三角引用结构中的施引行为倾向于包含多个功能性引用动机,而情感性引用动机则比较明确,多具有唯一性。根据本文对 9 442 条三角引用数据的动机标注结果,有近 31% 的 $B \rightarrow A$ 引用关系有两个及以上的功能性引用动机,而在一条引用关系中包含两个及以上情感性引用动机的情况仅占 14% 左右;同样地,在 $C \rightarrow A$ 引用关系中,同时包含两个及以上功能性引用动机的情况超过了 28%,而情感性动机仅有 9% 左右;在 $C \rightarrow B$ 引用关系中,同时含有两个及以上功能性引用动机的情况高达 35.4%,而情感性动机仅占 9.6%。因此,在作者的施引行为中,参考文献会起到多

种功能,但作者的施引情感一般是单方面的。
(2)对比在 18 种引用动机中 $B \rightarrow A$ 、 $C \rightarrow A$ 、 $C \rightarrow B$ 三种引用关系的分布情况,可以发现 $B \rightarrow A$ 与 $C \rightarrow A$ 两种引用关系的分布情况基本一致,而这两者的覆盖程度与另外一种 $C \rightarrow B$ 引用关系具有明显差别。一方面, $B \rightarrow A$ 与 $C \rightarrow A$ 两种引用关系均是文献 B、C 分别对同一篇文献 A 施加引用,此时,文献 B 与文献 C 的施引行为在引用主题、引用情境、引用方式上会存在相似,从而导致前两种引用关系的动机分布与文献 C 施引 B 存在差异。另一方面,3 种引用关系的动机分布差异也间接说明了在 $C \rightarrow A$ 中间接引用行为的存在。根据三角引用结构中文献的发表时间先后以及施引关

系发生的先后顺序,文献 B 在对 A 施加引用后,若文献 C 是通过文献 B 间接施引原始文献 A,那么文献 C 会在很大程度上参考、借鉴 B→A 的引用主题、引用情境、引用方式等,因而 B→A 与 C→A 的引用功能与引用情感分布基本一致。

此外,观察表 3 中“为了自引,使研究具有延续性而引用自己的文献”这一引用动机在 3 种引用关系中的分布情况,C→A 的自引与其他两种引用关系形成了明显的悬殊,这也从另一个角度反映了上述的间接引用现象。在三角引用结构中,文献 A 与文献 B、文献 B 与文献 C 两两之间多为传承性或相似性研究,因此,自引更倾向于发生在文献 A 与 B 或文献 B 与 C 之间,而在文献 A 与 C 之间难以产生作者自引关系,在原始文献 A 与追随文献 C 之间存在一种自引障碍和“鸿沟”,即三角引用结构中的间接引用现象。

(3)进一步详细对比三种引用关系在功能性引用动机中的分布情况。可以看到“引用参考文献说明现有文献未涉及该论题,阐明论题的新颖性或前沿性”“为综述、评论以前的文献而引用”这两种引用动机在 C→B 中的覆盖范围远高于在 B→A 与 C→A 中,而对于“引用参考文献的相关事实或数据说明该研究主题的重要性或必要性”“采用了参考文献中提出的新概念、新观点或新方法等”这些引用动机,B→A 与 C→A 的覆盖率要远高于 C→B。由此说明,在三角引用结构中,原始文献 A 是相关研究主题、领域或学科比较重要的文献,倾向于提供一些新的、开创性的思想、概念、观点或方法;而文献 B 则在引用关系中多起到文献综述、“中介桥梁”的作用。

(4)进一步详细对比 3 种引用关系在情感性引用动机中的覆盖范围。首先,可以看到在三角引用结构中,绝大部分引用情感为中性,相比之下带有情感倾向的正面引用动机要比负面引用动机多一些。因此,在三角引用结构中,施引文献的引用情感是隐藏的,通常会引用文献、观点或数据等不带明显情感倾向的内容,除中性引用之外,作者会更倾向于表达正面情感的引用,而进行质疑或否定等负面引用的情况相对较少。

其次,对比三角引用结构中 B→A、C→A、C→B3 种引用关系在 7 种情感性引用动机中的分布比例,可以发现在“对该领域的开拓者或创始人表示尊重”“对参考文献或其作者的工作表示肯定”这两种正面引用动机中,B→A 与 C→A 的覆盖比例远超 C→B,但在“用于文献综述”中,C→B 的分布比例远远高于 B→A 与 C→A。因此,文献 B、C 对原始文献 A 正面引用的高比例,说明文献 A 的学术价值与地位要高于三角引

用结构中的其他两种文献 B、C,在其所属学科和研究领域中一般具有较高的影响力和价值。此外,从引用情感动机角度,也可以得到与引用功能性动机一致的结论:文献 B 在三角引用关系中多用于文献综述和概括。

6 结论

三角引用现象是文献计量学中一种特殊的文献引用关系,其中蕴含着重要的“引用-被引用”内涵、机理、机制,具有较高的研究意义与应用拓展价值。本文在提出三角引用现象这一概念的基础上,首先通过整理已有关于引用动机的文献,总结出一套包含功能性引用动机和情感性引用动机的动机标注框架;其次,通过构建三角引用数据的获取步骤,从 CNKI 数据库中获取了 9 442 条三角引用关系数据;最后,由专业人员根据预设的动机框架,为每条三角引用数据中的 3 种引用关系标注相应的引用动机。实验结果发现:三角引用结构中的施引行为倾向于包含多个功能性引用动机,而情感性引用动机则多具有唯一性。另外,B→A 与 C→A 两种引用关系的引用动机分布基本一致,但与 C→B 存在差异。经过细粒度对比 3 种引用关系的这种差异,发现原始文献 A 是相关研究主题、领域或学科比较重要的、高影响力的文献,倾向于提供一些新颖、开创性的或基础的思想、概念、观点或方法,而文献 B 则在引用关系中多用于文献综述、概括。

本文的研究贡献在于:理论上,文献三角引用结构是直接引用、共被引与文献耦合的结合,有望在继承现有 3 种引用关系优越性能的基础上,进一步形成更加规范系统的一种新的引用结构,产生 1+1+1>3 的效果;并对各种引用关系和引用行为进行更加深入的理解和探索,如文献间的引用机制、学者特殊的引用行为、三角引用所反映的间接引用动机等。其次,本文总结了一套包含功能性引用动机和情感性引用动机的动机标注框架,该框架不仅适用于三角引用关系的动机标注实验,同时也可应用于其他引用关系的动机分类与标注研究中,并期望在应用中进一步优化、完善。

在实践贡献价值方面,本文发现了一种特殊的三角引用机制。B→A 与 C→A 的引用功能与引用情感分布基本一致,即原始文献 A 与追随文献 C 之间的“间接引用”机制,大量追随文献 C 的间接引用容易导致原始文献 A 的被引频次虚高,从而产生马太效应问题。因此,间接引用行为和马太效应都影响了引文数据的准确性,也削弱了引文分析作为学术评价工具的权威性和可信度,违背了引文分析、引文评价的本质和

意义。基于该发现,在引文评价中,我们可以考虑适当降低文献 A 所获被引频次的权重,以进行公平评价。其次,C→A 是在参考 B→A 的情况下发生,这种间接引用动机在一定程度上属于不规范引用问题,可以通过技术手段提前识别并排除这些引文相似度较高的不正当引用,以增强引文分析的可靠性,可为不当引用或规范性引用研究及其识别提供一个特殊的研究视角与思路。最后,在国家、相关机构组织等层面,应建立、健全各种类型科研文献、科研成果的全文存储与开放获取平台,尽可能为学者提供多语言、多种文献类型的文献资源;科研人员也应积极面对跨学科、跨语言等文献阅读带来的壁垒与困难,不断扩大在其他学科领域、研究方向的知识面,做到规范引用、合理引用。

除此之外,本研究还存在一些不足之处。首先,不同学科背景下的科研人员的引用行为可能存在着差异。在未来研究中,可以针对不同领域或学科科研人员的引用行为进行对比,找出可能存在的差异,从而提高模型的准确性,扩大模型的适用范围。

其次,学者在科研活动中的引用情境、态度、动机具有高度复杂性。无论是以上的文献外部特征分析,还是引文内容挖掘,都是建立在研究者的主观判断之上,极有可能与作者写作时的真实引用动机存在出入。因此,在未来工作中,有必要通过对研究人员进行问卷调查或访谈,从作者自身获得相对客观的经验感知数据。笔者将对具有一定发文量的学者进行半结构化访谈和问卷调查,探究三角引用中“间接引用”行为发生的概率、动机、影响因素等。并在此基础上,从引文生态学角度引入规避对策,构建间接引用的评价模型,以促进三角引用关系良性发展,在一定程度上解决计量学中负面的“间接引用”现象、“马太效应”问题等。

最后,传统的耦合关系基本上都是“不知情引用”,文献之间通过主题的关联性、相似性联系起来。但在耦合关系中也存在一种“潜在三角引用”的情形,即具有耦合关系的文献 B 和 C,文献 C 通过文献 B 了解到 B 中的参考文献 A,由于 A 是外文或更重要的文献等原因,文献 C 的作者在没有参阅 A 的情况下直接引用文献 A,但忽略了文献 B。这种引用现象较为复杂,由于文献 B、C 之间未发生真实的引用关系,很难从客观发生的文献数据集中发现这一现象。但这种“潜在三角引用”却是学者在科研工作中经常遇到的一类引用问题,同时也能为传统的文献耦合研究提供另外一种思考。因此在未来工作中,文献间的“潜在三角引用”及其动机研究是一个值得深入开展的问题。从文献三角引用视角出发,随着传统引文分析法与全

文本内容分析相结合,希望能够发现三角引用现象背后隐藏的科学家引用习惯与引用选择机制,并通过三角引用的文献影响力传递机制,为基于引文的学术评价方法提供改进思路,更好地为科学发展与科学评价工作服务。

参考文献:

- [1] KESSLER M. Bibliographic coupling between scientific papers [J]. American documentation, 1963, 14(1): 10-25.
- [2] SMALL H G. Co-citation in the scientific literature: a new measure of the relationship between two documents [J]. Journal of the American Society for Information Science, 1973, 24(4): 265-269.
- [3] 高楠,赵蕴华,彭鼎原. 基于引用关系与词汇分析法的研究前沿识别研究——以人工智能领域为例 [J]. 情报杂志, 2020, 39(4): 44-50, 13.
- [4] BO J. A comparison of two bibliometric methods for mapping of the research front [J]. Scientometrics, 2005, 65(2): 245-263.
- [5] BOYACK K W, KLAIVANS R. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? [J]. American Society for Information Science and Technology, 2010, 61(12): 2389-2404.
- [6] 刘宇,李武. 引文评价合法性研究——基于引文功能和引用动机研究的综合考察 [J]. 南京大学学报(哲学·人文科学·社会科学版), 2013, 50(6): 137-148, 157.
- [7] 刘茜,王健,王剑,等. 引文位置时序变化研究及其认知解释 [J]. 情报杂志, 2013, 32(5): 166-169, 184.
- [8] 祝小静. 团队型学科化服务模式实践与思考——以中国人民大学图书馆为例 [J]. 知识管理论坛, 2014(4): 22-26.
- [9] GARFIELD E. Citation index for science [J]. Science, 1955, 122(3159): 108-111.
- [10] MORAVCSIK M J, MURUGESAN P. Some results on the function and quality of citations [J]. Social studies of science, 1975, 5(1): 86-92.
- [11] VINKLER P. A quasi-quantitative citation model [J]. Scientometrics, 1987, 12(1): 47-72.
- [12] 朱大明. 参考文献的引用动机 [J]. 科技导报, 2013, 31(22): 84.
- [13] RONG T, MARTIN A S. Author-rated importance of cited references in biology and psychology publications [J]. Journal of documentation, 2008, 64(2): 246-272.
- [14] 史雅莉,赵童,杨思洛. 引证视角下科研用户的数据认知行为研究——基于扎根理论方法 [J]. 情报理论与实践, 2020, 43(6): 49-55.
- [15] 宋维翔. “王子”对“睡美人文献”引用的动机分析——基于邮件访谈调查的实证研究 [J]. 现代情报, 2018, 38(5): 32-36.
- [16] BROOKS T A. Private acts and public objects: an investigation of citer motivations [J]. Journal of the American Society for Information Science, 2010, 36(4): 223-229.
- [17] SINGSON M, SUNNY S, THIYAGARAJAN S, et al. Citation behavior of Pondicherry University faculty in digital environment: a survey [J]. Global knowledge memory and communication, 2020, 69(4/5): 363-375.
- [18] 张敏,刘盈,严炜炜. 科研工作者引文行为的影响因素及认知过

程——基于情感结果预期和绩效结果预期的双路径分析视角[J]. 图书馆杂志, 2018, 37(6): 74-84.

[19] 邱均平, 陈晓宇, 何文静. 科研人员论文引用动机及相互影响关系研究[J]. 图书情报工作, 2015, 59(9): 36-44.

[20] 丁文姚, 李健, 韩毅. 我国图书情报领域期刊论文的科学数据引用特征研究[J]. 图书情报工作, 2019, 63(22): 118-128.

[21] TEUFEL S, SIDDHARTHAN A, DAN T. Automatic classification of citation function[J]. Conference on empirical methods in natural language processing, 2006, 14(1): 103-110.

[22] 刘盛博, 丁堃, 张春博. 基于引用内容性质的引文评价研究[J]. 情报理论与实践, 2015, 38(3): 77-81.

[23] ROMAN M, SHAHID A, KHAN S, et al. Citation intent classification using word embedding [J]. IEEE Access, 2021, 9(1): 9982-9995.

[24] WANG M Y, ZHANG J Q, JIAO S J, et al. Important citation identification by exploiting the syntactic and contextual information of citations[J]. Scientometrics, 2020, 125(3): 2109-2129.

[25] GARFIELD E. Can citation indexing be automated? [J]. Essays of an information scientist, 1962(1): 84-90.

[26] 李卓, 赵梦圆, 柳嘉昊, 等. 基于引文内容的图书被引动机研究[J]. 图书与情报, 2019(3): 96-104.

[27] 马凤, 武夷山. 关于论文引用动机的问卷调查研究——以中国期刊研究界和情报学界为例[J]. 情报杂志, 2009, 28(6): 9-14, 8.

[28] CHANG Y W. A comparison of citation contexts between natural sciences and social sciences and humanities[J]. Scientometrics, 2013, 96(2): 535-553.

[29] 王文娟, 马建霞, 陈春, 等. 引文文本分类与实现方法研究综述[J]. 图书情报工作, 2016, 60(6): 118-127.

[30] OPPENHEIM C, RENN S P. Highly cited old papers and the reasons why they continue to be cited[J]. Journal of the American Society for Information Science, 1978, 29(5): 225-231.

[31] SEN S K. A theoretical glance at citation process[J]. International forum on information and documentation, 1990, 15(1): 1-7.

[32] BONZI S, SNYDER H W. Motivations for citation- a comparison of self citation and citation to others[J]. Scientometrics, 1991, 21(2): 245-254.

[33] 赖方中. 引文动机分析[J]. 四川警察学院学报, 2009, 21(6): 115-119.

[34] 陈晓丽. 引文类型比较分析[J]. 图书与情报, 1998(4): 3-5.

[35] 彭秋茹, 阎素兰, 黄水清. 基于全文本分析的引文指标研究——以 F1000 推荐论文为例[J]. 信息资源管理学报, 2019, 9(4): 82-88.

[36] VIERA A J, GARRETT J M. Understanding interobserver agreement: the kappa statistic[J]. Family medicine, 200, 37(5): 360-363.

作者贡献说明:

刘运梅: 负责研究设计与论文撰写;
张帅: 负责研究设计与论文修改;
司湘云: 负责研究设计与论文修改;
马费成: 负责研究设计, 指导与修改论文。

Investigation and Research on Triangular Citation Motivation
Based on Content Annotation

Liu Yunmei^{1, 2} Zhang Shuai^{1, 2} Si Xiangyun^{1, 2} Ma Feicheng^{1, 2}

¹ School of Information Management, Wuhan University, Wuhan 430072

² Wuhan University Information Resources Research Center, Wuhan 430072

Abstract: [Purpose/significance] Triangular citation in literature is a citation structure that integrates direct citation, co-citation and literature coupling. It has high research value and application expansion significance. In order to understand and dig out the triangular citation from a deeper level, tries to dig deeper into the hidden citation motivation through content annotation. [Method/process] Firstly, by sorting out the existing literature, summarized a set of motivational annotation framework that included functionality and emotion; secondly, constructed the steps for obtaining triangular citation data, and obtained 9 442 triangular citation relationship data in the CNKI database; finally, according to the preset motivation framework, professionals marked the citation motivation for each citation relationship in the triangular citation data. [Result/conclusion] The experimental results found that the citing behavior in the triangular citation structure tended to contain multiple functional citation motives, while the emotional citation motives were mostly unique; the distribution of the number of citation motivations for the two citation relationships of B→A and C→A were basically the same, but it was different from C→B; original literature A was a relatively important and high-impact document of related research topics, fields or disciplines, and tended to provide some new, pioneering and basic ideas, concepts, opinions or methods.

Keywords: triangular citation in literature citation motivation citation function citation emotion content annotation